DOCUMENT RESUME

ED 139 827 TM 006 320

AUTHOR	Donlon, Thomas F.
TITLE	A Practical Assessment of Field Dependence/Independence.
INSTITUTION	Educational Testing Service, Princeton, N.J.
SPONS AGENCY	Graduate Record Examinations Board, Princeton, N.J.,
PUB DATE	May 77
NOTE	35p.; Paper presented at the Annual Meeting of the New England Educational Research Organization (9th, Manchester/Bedford, New Hampshire, May 4-7, 1977)
EDRS PRICE	MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS	*Cognitive Style; *Cognitive Tests; College Students; Comparative Testing; Correlation; Higher Education; Item Analysis; Measurement Techniques; *Statistical Analysis; *Test Construction; Test Reliability; Test Validity
IDENTIFIERS	Embedded Figures Test; *Field Dependence Independence; *Figure Location Test (Donlon); Rod and Frame Test

ABSTRACT

This report summarizes research and development activity leading to the development of a new test of field-dependence/independence which was carried out with two related projects sponsored by the Graduate Record Examinations Board between September, 1974 and September, 1976. This new test is a group measure, which is machine scorable. In the first project, three versions of a machine scorable group measure, the Figure Location Test (FLT), were developed, using the same set of simple and complex figures. All three versions were administered to a sample of college students, along with the Group Embedded Figures Test and the Portable Rod and Frame Test. All three methods had acceptable levels of reliability and showed evidence of construct validity: however, the Direct-Machine readable version appeared to be most promising, based on analytical data and ease of administration. In a second study the machine readable Figure Location Test was reformulated into two 15-item sections, instead of three 10-item sections. Each of these formats was studied under three different time allowances. On the basis of these two studies, a version with 32 scorable items, organized into two 16-item sections, administrable in approximately 25 minutes, was developed. (Author/MV)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED139827

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

A Practical Assessment of Field Dependence/Independence

Thomas F. Donlon
Educational Testing Service
Princeton, NJ 08540

Ninth Conference of New England
Educational Research Organization
Manchester/Bedford, New Hampshire
May 4-7, 1977

TM006 320

Abstract

This report summarizes research and development activity leading to the development of a new test of field-dependence/independence. The activity was carried out with two related projects sponsored by the GRE Board between September, 1974 and September, 1976. This new test is a group measure, distinguished from related measures because it is machine scorable.

In the first project, three versions of a machine scorable group measure were developed and administered to a sample of college students, along with the Group Embedded Figures Test (GEFT) and the Portable Rod and Frame Test (PRFT). Each of these three versions used the same set of simple figures and complex figures and was referred to as the Figure Location Test (FLT). While all three methods had acceptable levels of reliability and showed evidence of construct validity, variant DM (The Direct-Machine readable version) appeared to be most promising, based on both the data considered and the ease of administration. When the item analytic data were examined, it appeared that additional developmental steps could be taken to further improve the psychometric properties of the Figure Location Test.

The second project undertook these recommended developmental steps. An additional 54 embedded figures items were developed, of which 45 were pretested on college students with known PRFT performance. In addition, the machine readable Figure Location Test studied in the first project was reformulated into two 15-item sections, in contrast to its original format of three 10-item sections. Each of these formats was studied under each of three time allowances: 24 seconds per item, 30 seconds per item and 36 seconds per item.

On the basis of these two studies, a version with 32 scorable items, organized into two 16-item sections, adminstrable in approximately 25 minutes, was developed for use in conjunction with additional research into the possible utility of measures of field-dependence/independence. The projected reliability for this version was .85 - .90.

A Practical Assessment of Field Dependence/Independence*

In recent years an increasing amount of research attention has been given to certain interesting constructs concerning mental performance which are referred to as cognitive styles. These cognitive styles have been conceptualized as typical modes of perceiving, remembering, thinking, and problem solving (Messick, 1969). They have been referred to as "styles" because they are thought to represent an individual's typical approach to cognitive tasks. Although a fairly diverse range of cognitive styles have been identified (Impulsiveness-restraint, Flexibility-rigidity) the most heavily researched cognitive style by far has been the field independence-dependence dimension.

Most of the basic research on the field-independence construct has utilized individually administered and scored tests. The three most commonly used measures are: (a) the Body Adjustment Test, in which the subject is seated in a tilted position in a room which is also tilted at some predetermined angle. The accuracy with which the subject can adjust the chair to the upright position, with the room remaining tilted, serves as the index of field independence; (b) the Rod and Frame Test (RFT), in which the subject is seated in a totally darkened room and must adjust to the upright a tilted luminous rod centered within a tilted luminous frame, while the frame remains in its initial position. A variation on this task is the Portable Rod and Frame Test (PRFT) which restricts the subjects perceptual field without a requirement for darkening the room. Again, accuracy is the scoring index used; and (c) the Embedded Figures Test(EFT) which requires the subject to locate a previously seen simple figure within a larger complex figure which has been organized so as to obscure the simple figure. Scores on the EFT are a function of the time taken to find the simple figure. The performance of subjects across these three tasks has been found (Witkin, Oltman, Raskin and Karp, 1971) to be relatively consistent. This finding is the basis for positioning the field independence construct as a perceptual style which reflects the tendency for a subject to be dominated by the overall organization of the surrounding field (field dependence) or to experience parts of the field as discrete from the surrounding context (Witkin, et al, 1971).

Witkin and his colleagues (1973, 1975) have made a case for the relevance of field independence-dependence to educational problems. Up until fairly recently, however, application of cognitive style research was limited because of the individual nature of the measurement procedures. With the development of the Group Embedded Figures Test (GEFT) by Witkin and his colleagues it became possible to obtain information about an individual's cognitive style in a group setting. The Group Embedded Figures Test consists of a series of complex figures within which a specific simple figure is located. The subject's task is to find and trace the simple figure directly in the test booklet. From the standpoint of a large-scale administration,

---

*Dr. Richard R. Reilly and J. David McKee, both formerly of ETS, cooperated in the work which is here reported.

however, the GEFT has the drawback of requiring trained personnel to score each item. Because of the need for methods of scoring which were both more rapid and more cost-effective, the present project was undertaken.

Development of Response Formats

One of the first possibilities considered was direct machine scoring of the traced simple figure. Under certain conditions such direct scoring would be technologically feasible. One way to make the simple figures readable, for example, would be to enlarge the figures to a size approaching that of one-half of the normal answer sheet. The embedded figure, when traced, would cover a detectable matrix of response locations. Another possibility would be to employ digitizers, such as those used for map-reading problems, which can electronically scan fields. The costs of these direct scoring approaches appeared too great to justify pursuing them any further, however.

The second approach considered was a quadrant approach used by Dees and his colleagues (Dees, 1968). This method breaks the complex figure into four equal areas or quadrants and requires the examinee to indicate the quadrant or quadrants in which the simple figure lies. This approach would have the advantage of employing a standard machine scorable answer sheet and, was considered promising because of results communicated by Dees (in a personal communication to Witkin) which reported a correlation of .80 between a 60-item test using the quadrant approach and the Embedded Figures Test, (N=262). The quadrant approach appeared less feasible, however, after a substantial number of new items had been developed. It became clear that the items used by Dees did not have the same kinds of properties as the items used by Witkin in the Embedded Figures Test and in the Group Embedded Figures Test. For example, one of the most significant differences between the Dees items and the Witkin items was the ratio of the size of the simple figures to the complex figure. For the Dees items this ratio was very small. That is, the simple figure was very small relative to the complex figure. For the Witkin items, however, the ratio of simple figure to complex figure was much larger. In general, Dees figures lacked a quality as Gestalten that Witkin and his associates value and strive to develop. Since the items developed for this project were more closely aligned with Witkin's point of view it was difficult to develop items which did not have the simple figure lying in at least three (and often four) of the quadrants. Because of the incompatibility between the type of item used in the EFT and GEFT and the Dees method the quadrant approach was abandoned.

The three response formats finally decided upon were: (1) the Hidden Letter (HL) technique--this method requires the examinees to actually trace the simple figure using a pencil or stylus. The pressure of the pencil or stylus copies the simple figure drawn onto a second page "capturing" a set of from one to four letters. The examinee then looks at the second page to see what letters have been enclosed by the simple figure and records these letters on the answer sheet; (2) the Overprinted Letter (OL) method--this approach presents a complex figure with a set of letters overprinted. The examinee's task is to find the simple figure and then record on an answer sheet the letters which lie within the simple figure; and (3) Direct Machine readable (DM) approach--this method presents the complex figures on a machine-readable answer sheet. The complex figures are faintly overprinted with bubbles. The subject is asked to find a simple figure and then to darken all the response locations (bubbles) which are contained in the simple figure. The test booklet itself is then separated into its component pages as machine-readable sheets. Unlike the other methods, there is no need for a separate answer sheet.

All three of these methods are parallel in terms of the location of the response possibilities. That is, the letters or bubbles appear at the same location on the complex figure for all methods. Each approach has certain advantages and disadvantages. The HL technique presents the complex figures without the possibly intrusive effect of overprinting. However, it requires costly paper, demanding accuracy in printing and booklet design, and a need to familiarize subjects with the technique for response. The OL method is a less demanding printing job but would seem to use maximally intrusive stimuli. It is also vulnerable to clerical errors of transcription as the subject moves from complex figure to answer sheet. The DM approach offers a possibly less intrusive set of stimuli, and may permit the subject to make more responses in a given time. Cost of scoring will be higher, however, because a separate answer sheet is not used.

Sample pages for each technique are attached as Appendix A. The appendix does not use pressure-sensitive paper to demonstrate the HL format, however.

Item Development and Pretesting - Phase I

The first step in the test development process was to generate a series of embedded figure items. Approximately 35 of these items were developed, each consisting of a complex figure with an associated simple figure embedded within.

The items were individually administered and timed. The order of presentation was varied, so that each item appeared once in each position. The effects of practice were assessed by plotting the

mean time taken to solve the problems in each ordinal position in which they appeared. The results of this plot are shown in Figure 1. It can be seen that the points form a fairly tight cluster and that the slope of regression line would be negative, suggesting that the later any item is presented in a series the easier it will be, other things being equal.

That there is a practice effect is not surprising. It is surprising, however, that the effect appears to be linear throughout the 35 items administered. A linear regression equation computed from the data presented in Figure 1 can be used to predict mean response time from ordinal position. The equation $y = 44.8 - .71x$ may be interpreted as indicative that for each ordinal position the mean response time is reduced by about 3/4 of a second. Three-fourths of a second would seem to be a fairly negligible effect in practical terms. It may be that this effect would level off above the maximum of 35 items considered in the present data.

Organization of Tests

The study was intended to contrast the measurement characteristics of the same content material in three versions, differing only in format. Each of these versions is called the FLT in the following text, interchangeably, where appropriate. Thus, a reference to "FLT total score," unless expanded by comment, is a reference to such a score secured from <u>any</u> version. The FLT was the new, machine-scorable item material in each case.

Each version of the FLT was divided into four separately timed sections. The first section consisted of five items administered with a two-minute time limit, and the remaining three sections each consisted of ten items administered with a five-minute time limit.

The FLT was administered to a volunteer sample of 307 students. Each examinee was paid $10 for participation in the study. Response format was spiraled so that approximately 100 examinees took each of the three versions. After taking a version of the new test, each examinee was administered the GEFT. Each examinee was also asked to provide his or her SAT Verbal and Quantitative scores and field of major study. In a session independent of the group administration each examinee also took the PRFT.

Results of Phase I

Although a total of 307 students were tested, a substantial number of these had missing data on one or more variables. As a result, the total number of usable cases was 252 (103 males and 149 females).

Figure 1

PRACTICE EFFECT

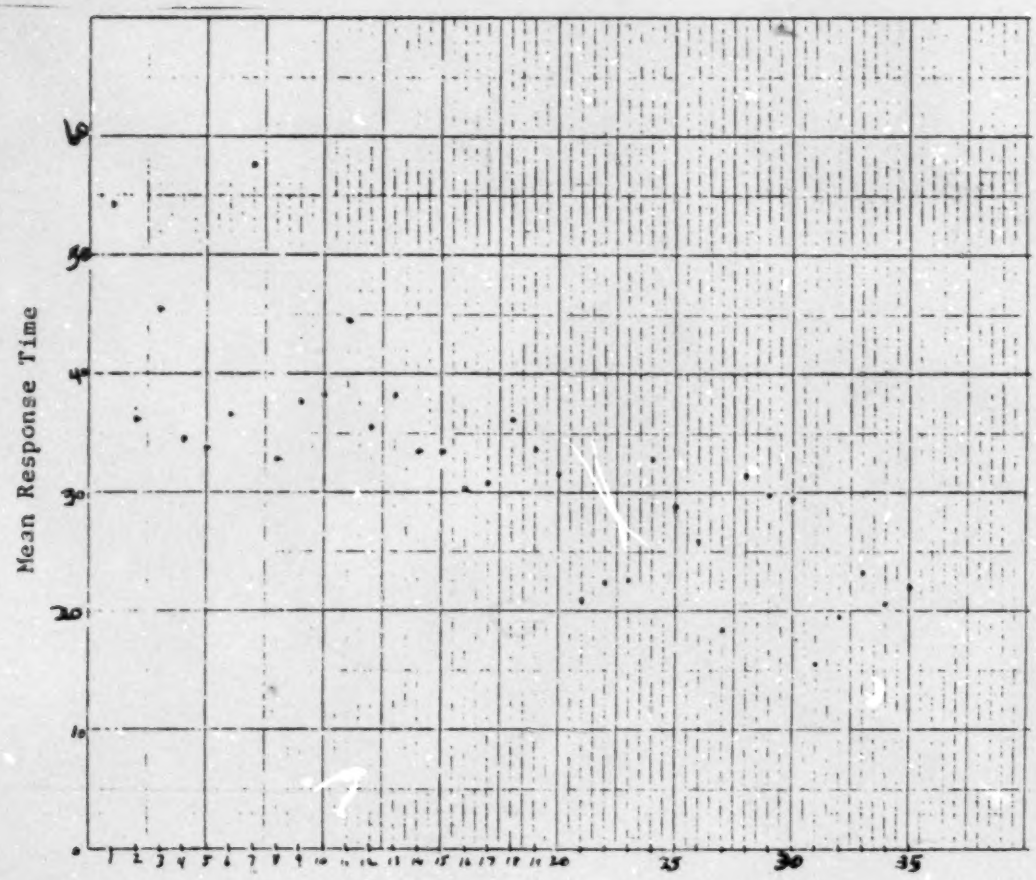


Table 1 shows means and standard deviations for each format-sex group on self-reported SAT-V and SAT-M, on GEFT and on PRFT. It is clear from the SAT means that the subjects are relatively high in academic ability. Although "national norms" are not available for GEFT and PRFT, the means for these two measures of cognitive style seem to indicate a relatively field independent group when compared with means obtained in other studies (high scores on GEFT and FLT indicate field independence, high scores on PRFT indicate field dependence).

Table 2 shows the reliabilities for the FLT in each response format and for each sex. All of the reliabilities are above .80 with a high of .91 for males using the DM format. The data of Table 2 suggest that any of the three response formats could be used to yield an acceptable reliability. Also shown in Table 2 are the reliabilities for the GEFT and PRFT for each subgroup. All reliabilities were computed using Cronbach's coefficient alpha. In the case of the GEFT and FLT the parallel subscores were derived from the separately timed sections. For the PRFT, the scores on the eight separate trials served as parallel subscores.

Results bearing on the convergent and discriminant validity of the FLT are presented in Table 3. A construct valid FLT would show relatively high relationships with GEFT and PRFT and relatively low relationships with SAT scores. Intercorrelations for FLT, PRFT, GEFT, and self-reported SAT-V and SAT-M are shown for each response format separately by sex and then pooled across formats within sex. Also, shown in parentheses are the correlations between each pair of cognitive style measures corrected for attenuation due to unreliability. In general, the picture appears more positive for the female data. For women all of the correlations between FLT and PRFT were significant, as were all those between FLT and GEFT. Correlations between FLT and SAT scores were somewhat high for the group using the HL format and for the group using the OL format.

For males, the correlations between FLT and PRFT in the group using the HL format was non-significant. Similarly, the FLT-PRFT correlation for males in the OL group was non-significant. In the DM group, however, the r between FLT and PRFT was .54, suggesting that, overall, the DM format worked best for both sexes.

The pooled results show a picture for FLT which is quite consistent with that shown for GEFT. The disattenuated correlations between GEFT and FLT were .70 for males and .76 for females. The correlations between FLT and other variables showed a pattern quite consistent with the pattern for the GEFT suggesting that the FLT closely parallels the GEFT.

Table 1

Means and Standard Deviations for PRFT, GEFT and SAT Scores by FLT Format Within Sex and Pooled Across FLT Format

	PRFT		GEFT		SAT-V		SAT-M	
	$\bar{x}$	S	$\bar{x}$	S	$\bar{x}$	S	$\bar{x}$	S
Male (HL)	28.06	23.06	15.18	2.68	558.30	72.60	579.45	109.20
Female (HL)	42.37	27.78	13.96	3.29	565.44	84.31	551.89	91.53
Male (OL)	41.81	43.23	14.38	3.70	592.97	83.49	632.76	67.22
Female (OL)	51.40	39.14	13.40	3.63	544.33	92.83	549.04	100.45
Male (DM)	32.43	29.51	13.89	4.79	552.91	101.70	587.53	120.97
Female (DM)	43.62	26.28	14.60	2.65	560.13	109.79	564.79	125.58
Pooled (Male)	33.90	32.85	14.47	3.86	567.43	87.55	598.89	103.88
Pooled (Female)	46.05	32.14	13.93	3.27	556.05	94.93	554.47	104.71

Table 2

Reliabilities for Measures of Field Independence Within Sex and Format and Pooled Across Format

	FLT			
	HL	OL	DM	Pooled
Males	.81	.81	.91	.84
Females	.81	.85	.82	.85

	GEFT			
	HL	OL	DM	Pooled
Males	.83	.79	.93	.87
Females	.80	.80	.81	.80

	PRFT			
	HL	OL	DM	Pooled
Males	.90	.94	.93	.93
Females	.83	.90	.79	.86

Table 3

Correlation Matrix for Males (N = 34)[a] and Females (N - 52)[a] Using HL Format

	FLT	PRFT	GEFT	SAT-V	SAT-M
FLT		-.13(-.15)	.52**(.63)	.17	.28
PRFT	-.50**(-.61)		-.16(-.19)	-.05	-.15
GEFT	.71**(.88)	-.43**(-.53)		-.15	.09
SAT-V	.44*	-.31*	.42**		.27
SAT-M	.57**	-.29*	.55**	.37**	

Correlation Matrix for Males (N = 32) and Females (N = 55) Using OL Format[a]

	FLT	PRFT	GEFT	SAT-V	SAT-M
FLT		-.17(-.19)	.63(.79)	-.11	.40*
PRFT	-.32**(-.36)		-.27(-.31)	-.17	-.28
GEFT	.60**(.73)	-.24*(-.28)		.20	.44**
SAT-V	.45**	-.16	.25*		.37*
SAT-M	.17	-.02	.40**	.20	

Correlation Matrix for Males (N = 37) and Females (N = 42) Using DM Format[a]

	FLT	PRFT	GEFT	SAT-V	SAT-M
FLT		-.54**(-.59)	.63**(.68)	.30*	.37*
PRFT	-.35*(-.43)		-.73**(.78)	-.33*	-.19
GEFT	.55**(.67)	-.44**(-.55)		.21	.25
SAT-V	.12	-.02	.05		.48**
SAT-M	.14	-.18	.20	.73**	

Table 3 (cont'd)

Correlation Matrix for Males (N = 103)
and Females (N = 149) Pooled Across Three FLT Formats

	FLT	PRFT	GEFT	SAT-V	SAT-M
FLT		-.35**(.40)	.60**(.70)	.21*	.35**
PRFT	-.48**(.56)		-.42**(.47)	-.15	-.14
GEFT	.63**(.76)	-.34**(.41)		.13	.23**
SAT-V	.38**	-.16*	.25**		.42**
SAT-M	.36	-.15*	.39**	.45**	

[a]Correlations for males are shown in upper diagonal. Correlations for females are shown in lower diagonal. Correlations in parentheses are corrected for attenuation.

*p > .05
**p > .01

Item Analysis Results

The role of errors in speed tests is a moot point in measurement theory. The individual items in the FLT obviously varied in intrinsic difficulty above and beyond the ability of a candidate to reach them. In evaluating the present version of the test it is clear that it is far from being "classically" speeded, in the sense of Gulliksen, where the candidates differ not at all in the accuracy of work. But this is a property which it shares with the Group Embedded Figures Test, a measure often judged as highly speeded. The GEFT for the present population was probably not all that speeded, yet its error variance contributed to a reasonably strong reliability and validity.

The complex role of speed in these measures may be indicated by the tendency of the more highly successful items to come from the middle of a section. If one divides Sections 2, 3 and 4 into three sets of items: first three, middle four, last three, the middle four seem to demonstrate an unusual proportion of successes by the criterion of correlation with PRFT. A crude measure of this is the simple average of 24 ("male" and "female") correlations for the 12 middle items versus the average for the 36 correlations for the first and last items. The "middle" item average is .20, the "extreme" item average is .14. There is no apparent figural characteristic of the material which would cause this phenomenon. While item variances could be a factor, with middle items more likely to be middle difficulty items, it raises the question of whether PRFT could not be better predicted by a score based on the sum of middle items exclusively. Clearly, items with supposedly greater speed variance, the last items in the various sections, are not the best predictors of PRFT.

This pattern of potential prediction from middle items is not observed when the criteria is the FLT score itself. There, the 24 "middle correlations average .493, the 36 "external" correlations average .494. Nor is it observed for GEFT. When the criterion is GEFT, the means are .301 "middle," .29 "extremes." To test the significance of the difference in the PRFT correlations, the Mann-Whitney test was applied. Table 4 shows the results

Table 4

Association Between Item Position and
Level of Item Correlation with PRFT

Phase I - Males

	Middle items	Extreme items	
Above median	9	2	11
Not above median	3	16	19
	12	18	30

χ^2 = 12.7

$\chi^2_{.995}$ = 7.9, 1 d. f.

Phase I - Females

	Middle items	Extreme items	
Above median	9	6	15
Not above median	3	12	18
	12	18	30

χ^2 = 5.0

$\chi^2_{.925}$ = 5.0

That the PRFT should seem to be better preeicted by middle items was an unanticipated finding mildly at variance with the expectations one would derive from the theory of speeded tests. It could, perhaps, be attributed to the somewhat complex position-related factors growing out of the test developer's practice of placing easy items first and the role of dropout in skewing item score distributions for late-appearing items. However, since it seemed potentially related to the validity of the FLT, a further investigation of it seemed indicated.

Speededness Data

The distributions of "Not Reached" scores were prepared separately for each section of the Figure Location Test for each of the six sex-by-format combinations. Table 5 presents the means and standard deviations.

Table 5

Means and Standard Deviations of Not Reached Score
Figure Location Test: Sections 2, 3, 4

		OL		HL		DM	
		M	F	M	F	M	F
Section 2	M	2.1	2.5	2.4	3.0	2.4	1.6
	σ	2.1	2.3	1.9	2.0	2.2	1.6
Section 3	M	2.3	2.5	3.2	2.7	1.9	1.5
	σ	2.1	2.3	2.4	2.4	2.2	1.7
Section 4	M	2.2	2.2	2.2	2.3	1.7	1.0
	σ	2.3	1.8	1.4	2.1	2.3	1.2

Summary of Phase I

The results of Phase I suggested that the FLT has a reasonably reliable test with a pattern of intercorrelations indicating a level of construct validity similar to the GEFT. The DM format was indicated to be the preferred format for continued study, since DM scores had the most positive pattern of relationships and the advantage of not requiring a separate answer sheet. This allowed the candidate more time per item and makes the overall examinee task less formidable than either of the other two methods.

The data suggested that the FLT might be improved in a number of ways: (a) determining whether a more optimal time/item ratio could be used with the DM format through further study of the appropriateness of time; (b) creating new items to raise the average level of item-criterion correlation which would improve the validity of the FLT; and (c) examining the role of item location effects in the variance of the FLT.

Phase II

The original 30 scorable items of the FLT, arranged in three 10-item sections, were restructured as two 15-item sections. This was done by retaining the ten items of original Sections 2 and

3 as the first ten items in the new Sections Two and Three and then assigning all odd numbered sections four items as 11 through 15 in Section Two, all even numbered sections four items as items 11 through 15 in Section Three. These two formats, called FLT1 (the original, three-section format) and FLT2 (the new two-section format) were each administered under three separately times conditions.

Time-per-item allotments both more generous (36 seconds/item) and less generous (24 seconds/item) than the original 30 seconds per item were selected. These were chosen as reasonably distinct from the original format and as leading to total-time-per-section requirements which were stated in whole minutes (with one exception). This was a decided administrative convenience to the already well-burdened test administrators. The design thus yielded a two-factor time x format matrix in a 2 x 3 configuration. A strong attempt was made to randomly assign subjects to sessions. However, the participants, being students, had to meet the demands of their class schedule as well. This inevitably led to certain modifications.

Correlational Results Phase II

Table 6 shows the reliability for each of the principal measures FLT, GEFT and PRFT for each of the six experimental subgroups. All reliabilities were computed as coefficient alpha. For the FLT mrasures, there were variously two or three subsections in the sum of section variances. FLT1 had three score-contributive subsections, FLT2 had two. GEFT coefficients considered the two score-contributive sections of this list. The PRFT reliabilities are based on the use of a split-half approach, using the sum of the odd trials and the sum of the even trials, rather than all eight trials as independent elements, as in the corresponding calculations in Phase I.

The results for FLT are logically attributable to factors in addition to sampling variation. Each of the six reliabilities for FLT represents a unique combination of format and time. Conclusions regarding these factors as determiners of reliabilities, however, are rather difficult to make.

The three coefficients for FLT2 are each greater than any for FLT1, a pattern that does not hold for the other two measures. The variability in these indices is quite great, however, and the interpretation must be guarded. No clear impact of timing variations on reliability is demonstrated.

Table 6

Reliability[2] of FLT, GEFT and PRFT
in Each of Six Subgroups

Subgroup Form	Time	N	FLT	GEFT	PRFT
FLT1	24 s/i	80	.80	.84	.90
FLT1	30 s/i	70	.68	.88	.62
FLT1	36 s/i	66	.77	.68	.72
FLT2	24 s/i	58	.84	.68	.70
FLT2	30 s/i	72	.84	.86	.84
FLT2	36 s/i	78	.86	.78	.86

[2]Reliability computed as α

Table 7 shows the intercorrelations among FLT, PRFT, GEFT, SAT-V and SAT-M for each sex in each of the six experimental conditions. The effect of time limits on these correlations, as gauged by the criterion of statistically significant outcomes, seem to favor the 30 seconds/items timing. If one focusses only on the correlations of FLT with PRFT and GEFT, then the pattern in Table 8 emerges. While the correlations with GEFT were highly significant in each case, only the 30 second time condition yielded consistently significant correlation with PRFT. This it did for both males and females. The patterns of correlation within SAT-V and SAT-M were complex. Males showed a generally greater tendency toward higher correlations between FLT and with SAT-V and SAT-M. The median correlation for males under the six experimental conditions was .48 for both SAT-V and SAT-M. For females, it was .28 and .21, respectively. Of the twelve FLT-SAT correlations for males, ten were significant; for females, only four. The results were somewhat similar to those exhibited in the Phase I sample, where males had shown consistently higher (SAT-M)--(FLT) correlations as opposed to (SAT-V)--(FLT).

Table 9 shows the correlations among the FLT, GEFT and PRFT for each experimental condition for the sexes combined. The most consistently significant time condition is the 30 seconds per item.

Table 7

FLT1 Correlations for Males (N = 30)[a] and Females (N = 47)[a] using 24 Second Time

	FLT	PRFT	GEFT	SAT-V	SAT-M
FLT		-.33	.78**	.62**	.73**
PRFT	-.18		-.46*	-.28	-.27
GEFT	.62**	-.25		.56**	.69**
SAT-V	.34*	-.24	.22		.80**
SAT-M	.27	-.25	.20	.72**	

FLT1 Correlations for Males (N = 39) and Females (N = 30) using 30 Second Time

	FLT	PRFT	GEFT	SAT-V	SAT-M
FLT		-.38*	.76**	.57**	.70**
PRFT	-.44*		-.54**	-.30	-.32
GEFT	.58**	-.57**		.49**	.58**
SAT-V	.32	-.24	.29		.60**
SAT-M	.42**	-.34	.55**	.74**	

FLT1 Correlations for Males (N = 28) and Females (N = 35) using 36 Second Time

	FLT	PRFT	GEFT	SAT-V	SAT-M
FLT		-.21	.70**	.23	.43*
PRFT	-.05		-.17	.01	-.11
GEFT	.65**	-.28		.30	.48**
SAT-V	.23	-.15	.24		.81**
SAT-M	.04	-.41*	-.01	.63**	

[a]FLT1 correlations for males are shown in upper diagonal. Correlations for females are shown in lower diagonal.

Table 7 (cont'd)

FLT2 Correlations for Males (N = 31) and Females (N = 21) using 24 Second Time

	FLT	PRFT	GEFT	SAT-V	SAT-M
FLT		-.27	.70**	.82**	.46*
PRFT	-.31		-.55**	-.38*	.37*
GEFT	.63**	-.29		-.67**	.32
SAT-V	-.08	-.33	.12		.61**
SAT-M	-.12	.30	.36	.36	

FLT2 Correlations for Males (N = 31) and Females (N = 39) using 30 Second Time

	FLT	PRFT	GEFT	SAT-V	SAT-M
FLT		-.56**	.72**	.37*	.20
PRFT	-.51**		-.46**	-.14	-.20
GEFT	.69**	-.51**		.43*	.23
SAT-V	.39*	-.16	.27		.47**
SAT-M	.15	-.08	.09	.59**	

FLT2 Correlations for Males (N = 36) and Females (N = 42) using 36 Second Time

	FLT	PRFT	GEFT	SAT-V	SAT-M
FLT		-.27*	.63**	.38*	.51**
PRFT	-.32*		-.34*	-.19	-.41*
GEFT	.58**	-.36*		.31	.49**
SAT-V	.16	.11	.10		.73**
SAT-M	.34*	.04	.28	.82**	

*$p < .05$

**$p < .01$

Table 8

Level of Significance for Correlation of FLT1 and FLT2 with PRFT and GEFT by Sex and Timing Condition

		PRFT		GEFT	
Time		M	F	M	F
24	FLT1	N.S.	N.S.	.01	.01
	FLT2	N.S.	N.S.	.01	.01
30	FLT1	.05	.05	.01	.01
	FLT2	.01	.01	.01	.01
36	FLT1	N.S.	N.S.	.01	.01
	FLT2	.05	.05	.01	.01

Table 9 shows the correlations among the FLT, GEFT and PRFT for each experimental condition for the sexes combined. The most consistently significant time condition is the 30 seconds per item. In general, the FLT2 format gave higher values for the FLT-PRFT correlations. Only in the FLT2--30 second item combination did the FLT show a relationship with PRFT as great as that shown by GEFT.

Dropout

Table 10 shows the average and standard deviation of the "Not Reached" statistic, which is defined as a nonresponse to an item which is not followed by any response to a later item. There are clear regularities in the effect time condition, although these effects are more marked for FLT2, with its longer sections. When these data are compared to those of Table 5, the general rate of work estimates for DM-variant material is sustained. Each section in the test tends to show less dropout than its predecessors, indicating the gradual warm-up effect that Witkin has reported and that was evident in the individual item tryouts of Phase I. Levels for the FLT2 format, with its additional item, are markedly greater for the more stringent time limits. Since the test is only semi-speeded, and since speed variance is not directly associated with the criterion measure PRFT, no firm evaluations can be placed on the dropout data. The tests seem moderately and suitably speeded.

Table 9

Intercorrelations Among GEFT, FLT and PRFT
for Six Experimental Conditions for Combined Sexes

Format	Time	FLT-GEFT	FLT-PRFT	GEFT-PRFT	N[†]
FLT1	24 sec/item	.70**	-.24*	-.36**	78
FLT1	30 sec/item	.68**	-.39**	-.53**	69
FLT1	36 sec/item	.68**	-.11	-.24	64
FLT2	24 sec/item	.67**	-.29*	-.46**	57
FLT2	30 sec/item	.70**	-.53**	-.48**	71
FLT2	36 sec/item	.61**	-.29*	-.38**	78

[†]Ns may vary by one or two cases for specific coefficient.

*$p < .05$

**$p < .01$

Table 10

Means and Standard Deviations of the
Not Reached Scores for Each Section of
FLT for Each Time Condition, Combined Sexes

Form	Time Condition (See/item)		Section 2	3	4
FLT1	24	m	2.5	2.1	2.1
		σ	1.9	1.8	2.5
FLT1	30	m	1.9	1.2	0.9
		σ	2.0	1.8	1.4
FLT1	36	m	1.2	1.2	1.0
		σ	2.0	1.9	1.7
FLT2	24	m	4.9	3.8	---
		σ	3.1	3.2	---
FLT2	30	m	2.6	2.5	---
		σ	2.6	2.5	---
FLT2	36	m	1.3	1.3	---
		σ	1.8	1.9	---

Item Correlation Pattern

The phenomenon of higher validity for so-called middle items was an interesting feature of Phase I. The data processing schedule for the results of Phase II was such that a study of the item validities of FLT1 and FLT2 items against the PRFT criterion could not be accomplished. However, such data were examined for a 45-item pretest section administered concurrently, FLTX.

FLTX materials were organized into three sections of fifteen items each. On the basis of the 3-4-3 pattern which was used to demonstrate "end" and "middle" items in Phase I, a 4-7-4 pattern was chosen for Phase II. Each item was characterized as equal to or exceeding the median in the distribution of item criterion correlations with GEFT, PRFT and FLTX-Total.

The results for GEFT and PRFT yielded equivalent two-way contingency tables, presented in Table 11. The pattern is significant at the .05 level. FLTX items were not organized by difficulty level with the same preliminary pretesting that was used in FLT1. While later items are doubtless more difficult, and more subject to drop-out, these statistical factors seem likely to effect correlation with FLTX total as with PRFT. As in Phase I, however, the correlations with FLT Total do not confirm the pattern.

General Implications for FLT Organization

The evidence on the effect of test length and time limit is not unequivocal. The maximum reliability estimate, taken as coefficient alpha, is .86 for a test organized as two 15-item sections. In general, reliability in this format exceeded the corresponding figure for the test when organized as three 10-item sections.

When correlations with criteria are examined, twelve groups are available in the 2x3x2 design (test organization x timing variation x sex). They are quite small, ranging from 28 to 47 cases. Perhaps the most critical values are the correlations with PRFT. The highest correlations for both males and females are secured with the 30-second time limit and the 15-item sections. With the exceptions of males on FLT1, values on FLT2 are uniformly greater than the corresponding FLT1 values. All four of the correlations based on 30-second timing exceed any of the other values.

When GEFT is considered as a potential criterion, 30-second timing is superior for an FLT2 organization, but the generally stronger correlations between FLT1 and GEFT for males give this format an overall advantage. Similarly, the patterns for correlation between FLT and SAT-V and SAT-M are complicated. Generally lower correlations are observed with SAT-M for FLT2, the 15-item format, but no clear superiority is demonstrated.

If the level of significance alone is considered, Table 18 indicates that the strongest associations with PRFT are for FLT2

Table 11

Association Between Item Position and Level of Item Correlation with Three Criteria Phase II - Total Sample

PRFT and GEFT

(These Variables Yield the Same Total)

	< mdn	> mdn	
"Middle" Item	6	15	21
"End" Items	16	8	24
	22	23	45

$\chi^2 = 5.1$
$p < .05$, 1 d.f

FLTX Total

	< mdn	> mdn	
"Middle" Items	8	13	21
"End" Items	14	10	24
	22	23	45

$\chi^2 = 1.8$ n.s.

format at 30-sec/item. For pooled sexes, as in Table 9, the highest correlations are found for the FLT2 - 30-sec/item combination.

Finally, the dropout data offer evidence that the Not Reached characteristics are more marked for FLT2 for the different variations in timing. In balance, over all of the characteristics, the properties of the FLT2 organization are superior to those of the FLT1 organization. It is impossible to attribute this to the influence of an expanded "middle items" phenomenon, but this phenomenon was confirmed in the Phase II data.

Summary

The major results of these studies are as follows:

1. The successful development of a machine-scorable measure of field dependence-independence.

2. Demonstration of reasonable levels of congruent and divergent validities in the patterns of correlation with GEFT, PRFT, SAT-V and SAT-M.

3. Demonstration that the 30 seconds/item time condition offers slightly superior psychometric properties.

4. Demonstration that the placement or organization of items into longer item formats, involving 15-items per section, is superior to other placements in validity defined as correlation with PRFT.

5. Demonstration that a variety of items can be produced and adapted to the new method; that these are psychologically acceptable to candidates.

6. Demonstration of a position effect governing item-criterion correlations in a manner not predicted from the theory of speeded tests.

References

Dees, J. W. Personal communication to H. Witkin, 1968.

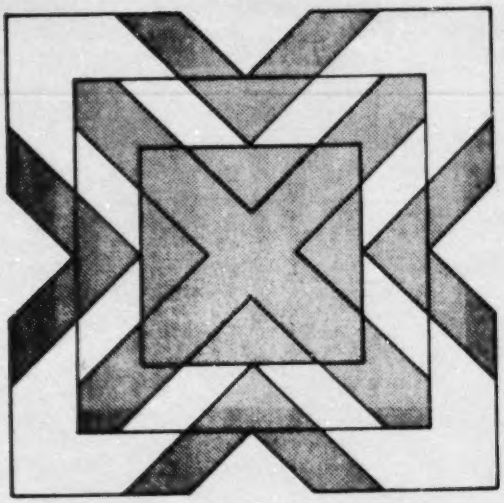
Messick, S. The criterion problem in the evaluation of instruction. Assessing possible, not just intended outcomes. Research Bulletin 69-89, Princeton, N.J. Educational Testing Service, 1969.

Witkin, H. A., Oltman, P. K., Raskin, E. and Karp, S. A. A Manual for the Embedded Figures Test. Palo Alto, Calif.: Consulting Psychologists Press, 1971.

Witkin, H. A. A cognitive-style perspective on evaluation and guidance. Proceedings of the 1973 Invitational Conference on Testing Problems--Measurement for Self-Understanding and Personal Development. Princeton, N.J.: Educational Testing Service, 1974, 21-27.

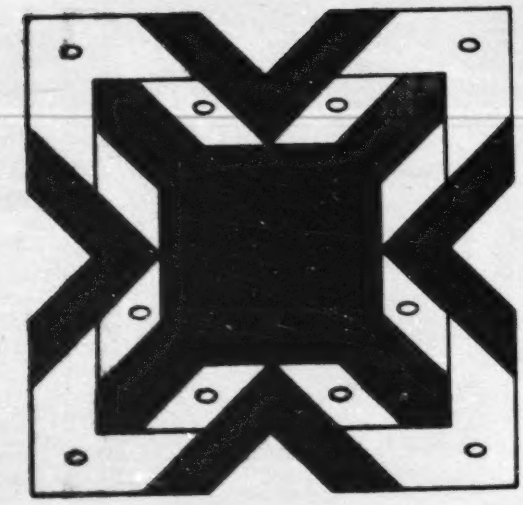
Witkin, H. A., Moore, C. A., Goodenough, D. R. and Cox, P. W. Field-dependent and field-independent cognitive styles and their educational implications. Research Bulletin 75-24. Princeton, N.J.: Educational Testing Service, 1975.

APPENDIX A

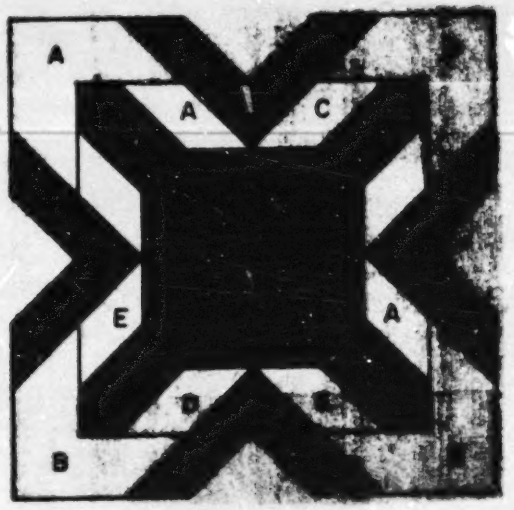


HL Method

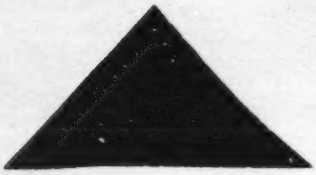
See Next Page For Hidden Letters



DM Method

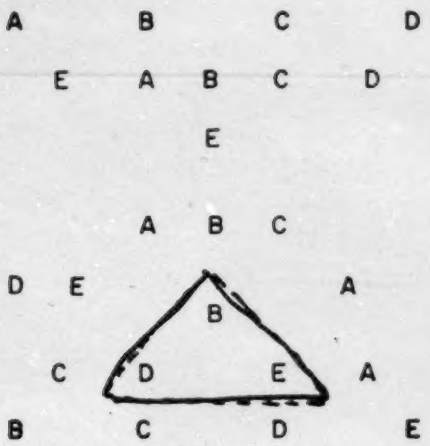


OL Method



The Associated Simple Figure

APPENDIX A (cont'd)



HL Hidden Letters, with
Sample of Outline
Produced by Tracing.

Subject Answers BDE

APPENDIX B

Notes on the Development of FLT Items

Certain principles have to be observed so that the items will be as valid as possible, machine-scorable, and written quickly. Items, therefore, have to be somewhat restricted in size and complexity, but otherwise creativity is unimpeded.

The authors of the FLT observed three approaches to their own item writing: 1) the complex figure can be drawn first, the simple figure selected, then shading can be added to hide the simple figure; 2) the simple figure can be drawn first then expanded until it forms a complex design; 3) an existing item or rejected one can be modified until acceptable.

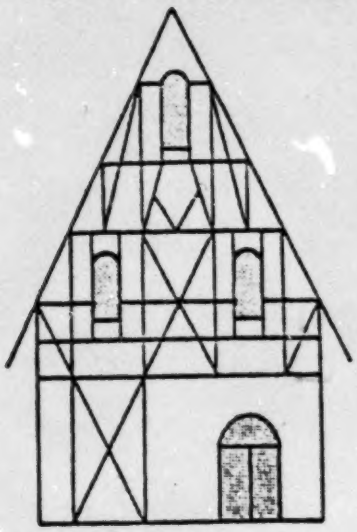
The validity of the items depends first upon the test-taker's tendency to form a gestalt--a sense of a whole design--upon viewing the larger complex figure. Second, he or she looks at a simple figure and then tries to find it in the complex one. This requires that the gestalt be broken down mentally in order to see the simple figure. The person somehow has to overcome the compelling design and to focus only on the simple figure that he or she wishes to see. It is the job of the item writer to make this task as hard as possible.

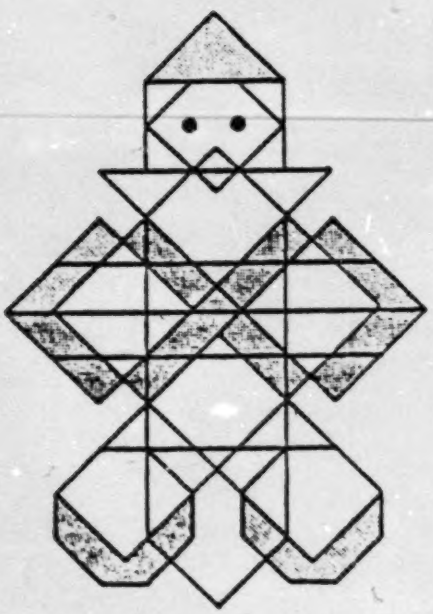
Symmetry, for example, is identified as a property of the complex figure that gives it a gestalt-like quality. The item writer must be aware, however, that symmetry in the figure may produce a double solution or it may require that the simple figure lie in the very center. These relationships between symmetry in the simple and/or the complex figure are exemplified so that the item writer can use symmetry to his or her advantage and guard against the problems it creates.

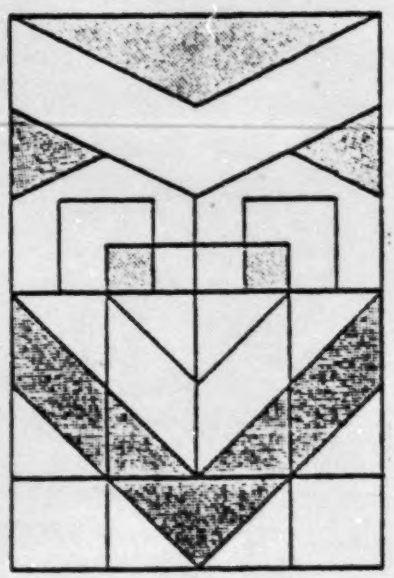
Shading, heavy lines, 3-D effects, and illusions are illustrated as effective methods of hiding the simple figure as well improving a gestalt. Illustrated below are some of the new types appearing in the FLTX.

APPENDIX B (cont'd)

1. Real-life figures: Some of the complex figures so closely resembled actual objects or faces that they can be referred to as "real-life" figures. The "half-timbered house," the "robot," and the "shield" are illustrated below:



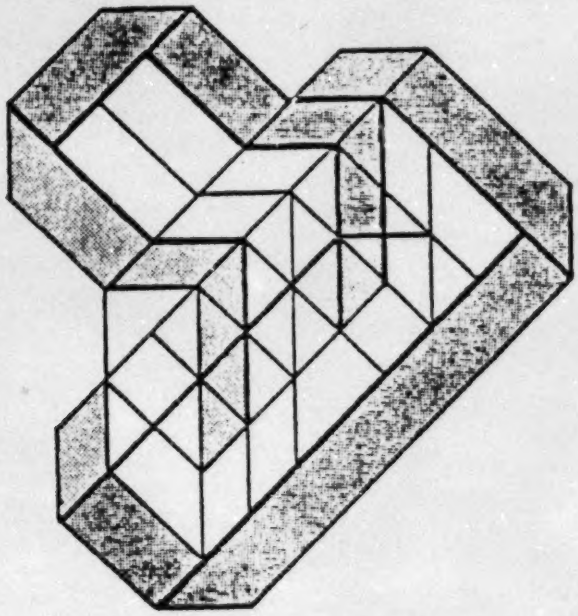




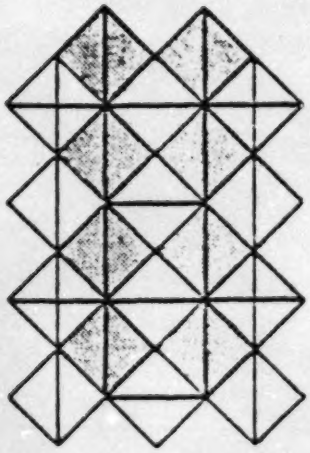
These items were used because the figures, insofar as they resembled objects, were judged to better-than-average gestalts.

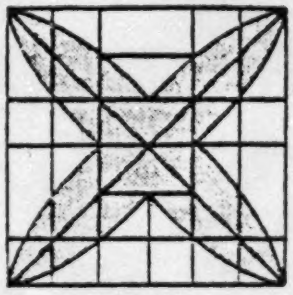
APPENDIX B (cont'd)

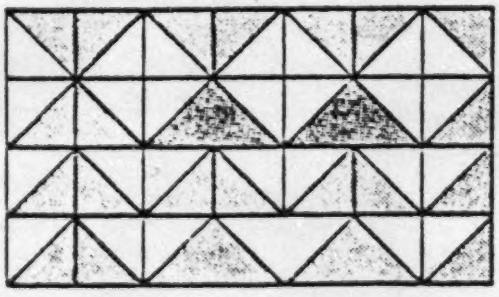
2. **Folding strips:** One way of creating a gestalt was to make the complex figure appear as if it was one or more continuous folding and weaving to form a pattern. Use of shading helped to produce this effect. The 'strip' was designed to break up the simple figure. Four items of this type were used in the FLTX; one is shown below:



3. **The tapestry design:** Three complex figures resembled a woven tapestry such as those shown below:

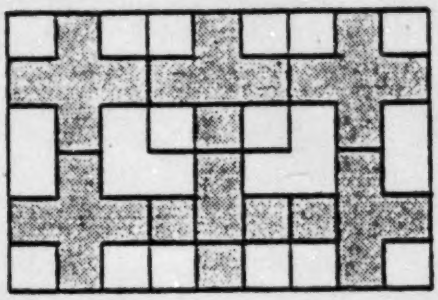




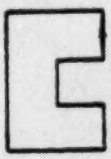


APPENDIX B (cont'd)

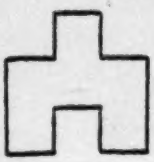
4. The H-figure (and variations): It was discovered that one simple figure was shaped like the letter H, and the complex figure vaguely resembled a maze, crossword puzzle, or tapestry. One such item is shown below:



There were five H-items used in the FLTX and three variations on it in which the simple figure was either



or



In the construction of the FLTX, the more difficult items were generally placed at the ends of each timed section. Among the difficult items were the H-figures. Whether these items correlated poorly with the PRFT because of their placement or because of some other property of the item cannot be determined from these data alone. None of the patterns based on item type as defined by these catagories could explain why some items were successful and others were not, except possibly the H-items. However, these were too few in number to say with any confidence that they were invalid.